Trustwave®
Smart security on demand

TRUSTWAVE SEG SPAMCENSOR EXPLAINED

Table of Contents

About This Document 2
1 SpamCensor Defined 3
2 How Does SpamCensor Work? 3
3 How Are the Rules Made and Scored? 3
4 Why Isn't SpamCensor Updated As Often As a Virus Scanner? 4
5 How Effective is SpamCensor? 4
6 More Information 5
About Trustwave 6

About This Document

SpamCensor is one of the core anti-spam technologies in Trustwave SEG. Over many years it has proven to be an effective, flexible and extensible tool for fighting spam.

This document addresses the following common questions:

- What exactly is SpamCensor?
- How does it work?
- How are the rules made and scored?
- Why isn't it updated as often as anti-virus tools?
- How effective is it?

1 SpamCensor Defined

Technically, SpamCensor is a special Trustwave SEG category script. Category scripts use XML script files to provide flexible multi-faceted content scanning.

SpamCensor is supplied and automatically updated by Trustwave. It is currently built and published multiple times per week by anti-spam experts in Trustwave's SpiderLabs Email Security team.

SpamCensor is a key element in the Trustwave SEG anti-spam toolkit. While it is a powerful tool, it's important to realize it is one element in the Trustwave SEG anti-spam solution. SpamCensor happily works in conjunction with other features such as SpamProfiler, Reputation Service lookups, URL filtering, Directory Harvest Attack (DHA) prevention, receiver block rules and anti-spoofing. This document only discusses SpamCensor.

2 How Does SpamCensor Work?

SpamCensor is a heuristic filter that consists of thousands of tests or rules for spam. It is not a signature-based system like an anti-virus scanner, where each signature operates independent of the other signatures. Rather it is a scoring-based system where rules work in combination, to end up with a total score which represents the 'spamminess' of a message.

Many different types of tests or rules are applied against each message. The header is checked for irregular fields and data, including the telltale footprints of the spammers' tools. The message body is checked for the characteristic patterns of key words and phrases and their relationship to other words and phrases. HTML code is parsed, again searching for spammers' footprints. The message, and its components, size and structure are also checked.

When the threshold of 60 is reached, the message is considered spam and action can be taken. The following message log illustrates this:

```
------ Category <Spam> evaluation result ------
- EV_GENR_XSmallBodyLT100: (24.00) Msg body is between 5 and 100 bytes
- HTTP_LINK: (1.00) HTTP link in message
- MSG_SIZE_1: (-1.00) IsBigger than 1K
- MT_GENR_XSmallHyperlnk: (10.00) Message body less than 100b with web link
- RH_GENR_HeaderSus21: (20.00) Header pattern common in spam
- RH_GENR_HeaderSus215: (15.00) Header pattern common in spam
- RH_GENR_HeaderSus247: (10.00) Header pattern common in spam
- RH_GENR_HeaderSus49: (22.00) Header pattern common in spam
- RH_GENR_HeloIP: (5.00) IP used in any HELO - current or previous
- RH_GENR_HeloIPtoMM: (25.00) IP used in HELO to MailMarshal
- RH_GENR_Localhost: (20.00) Received from localhost
- RH_GENR_XMailerMissing: (1.00) Missing X-Mailer: Field
- TC_HTTP_SEX: (16.00) Http link contains sex
Total score: (168.0) required(60.0)
SpamFilter: Version 130 26-September-2006
 ------ End of Category <Spam> evaluation ------
```

3 How Are the Rules Made and Scored?

SpamCensor's rules are crafted by members of the Trustwave SpiderLabs Email Security team. The team is a group of security analysts who examine live spam streams for patterns and trends. The experience of

the team and knowledge of spam patterns or 'traits' is a key factor in the effectiveness of the SpamCensor.

This 'human factor' should not be undervalued. Sometimes in the anti-spam community a lot of weight is given to automatic machine-learning systems, where human input is minimal. These systems serve a purpose, but can easily create problems if not configured, trained and monitored correctly.

Trustwave uses a number of back-end, proprietary tools that machine-learn by applying statistical analysis techniques to large volumes of spam and legitimate emails. The tools are especially good at highlighting patterns and trends that are not easily seen by humans. The results are fed into SpamCensor, but only after being validated by the SpiderLabs experts.

The values assigned to each SpamCensor rule are generated by an automated scoring algorithm. Each SpamCensor release is trained on over a million messages to optimize the filter. This maximizes spam detection and minimizes false positives. Fresh data is added daily to keep the training program up to date. SpamCensor is then delivered to the customer, updated and pre-trained, with no extra configuration needed.

4 Why Isn't SpamCensor Updated As Often As a Virus Scanner?

Sometimes customers ask why the SpamCensor is not updated as often as other anti-spam solutions or their virus scanner. The simple answer is that it is not a signature-based system. In such systems, each signature is an independent entity and equates to a single message. In the SpamCensor, rules are *heuristic* and *interdependent* which means individual signature updates are unnecessary. A longer release cycle is used to carefully craft, train, and test each SpamCensor prior to release.

This approach to spam detection results in a highly *predictive* filter. The bulk of spam is sent by relatively few spammers so patterns seen today are often repeated tomorrow. The readable content of a spam message may differ, but how the message is put together doesn't. The ability of SpamCensor to detect these underlying trends is what makes it so effective.

5 How Effective is SpamCensor?

The SpiderLabs team closely monitors the performance of the SpamCensor on Trustwave SEG servers, which are subjected to live streams of incoming spam. SpamCensor's detection rate, by itself, is typically around 99%.

When SpamCensor is combined with other recommended rules (such as IP Reputation Services, SpamProfiler, and URLCensor) the total effectiveness of Trustwave SEG against spam can exceed 99.7%.

6 More Information

This document provides basic information about the Trustwave SEG SpamCensor technology. With its unique heuristic capability, and backed by the resources of the Trustwave SpiderLabs team, the SpamCensor has proven to be a robust and highly effective anti-spam filter that is easily deployed.

To learn more about SpamCensor best practices, refer to the Trustwave SEG Anti-Spam and Anti-Malware Basics document, available from the SEG Support pages on the Trustwave website.

About Trustwave

Trustwave is a leading provider of compliance, Web, application, network and data security solutions delivered through the cloud, managed security services, software and appliances. For organizations faced with today's challenging data security and compliance environment, Trustwave provides a unique approach with comprehensive solutions that include its TrustKeeper® portal and other proprietary security solutions. Trustwave has helped hundreds of thousands of organizations—ranging from Fortune 500 businesses and large financial institutions to small and medium-sized retailers—manage compliance and secure their network infrastructures, data communications and critical information assets. Trustwave is headquartered in Chicago with offices worldwide. For more information, visit https://www.trustwave.com.